

OCR A Level

Computer
Science

H446 – Paper 1

3

Relational databases and normalisation

Unit 4

Exchanging data



PG ONLINE

Objectives

- Explain the concept of a relational database
- Normalise relations to third normal form
- Understand why databases are normalised

Relational database design

- In a relational database, data is held in tables, also known as **relations**
- One row in the table holds one record
- Each column represents one attribute
- Each relation should hold data about a single **entity**

Relational database design

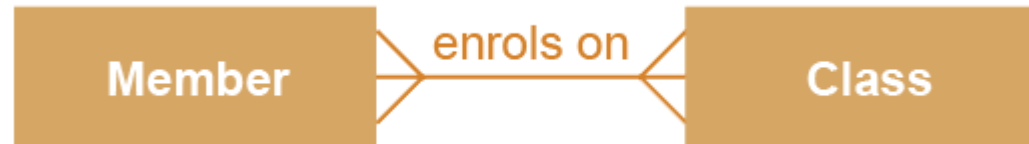
- Example: A database holds details of gym members and the classes they take
- A first attempt at designing the database might use a **flat file**, holding all the data in one table

MemberID	MSurname	MFirstName	ClassID	Cname	Day
M12	Sharpe	Norman	C100 C103 C114	Pilates Zumba Yoga	Monday Monday Wed
M13	Bunn	Martha	C100 C115	Pilates Cycling	Monday Wednesday
M16	Jones	Polly	C103	Zumba	Monday
M18	Jolly	Brian	C114	Yoga	Wed



Relational database design

- Relationship between gym members and the classes they take is many-to-many



MemberID	MSurname	MFirstName	ClassID	Cname	Day
M12	Sharpe	Norman	C100 C103 C114	Pilates Zumba Yoga	Monday Monday Wed
M13	Bunn	Martha	C100 C115	Pilates Cycling	Monday Wednesday
M16	Jones	Polly	C103	Zumba	Monday
M18	Jolly	Brian	C114	Yoga	Wed



Normalisation

- Normalisation is a process used to come up with the best possible design for a database
- Tables should be organised so that data is not duplicated in the same table or in different tables
- The structure should allow complex queries to be made
- There are three stages in normalisation, called First Normal Form (1NF), Second Normal Form (2NF) and Third Normal Form (3NF)

First normal form (1NF)

- A table is in first normal form if it contains no repeating attributes or groups of attributes
- All attributes must be atomic – a single attribute cannot consist of two data items such as firstname and surname
- This would make it difficult or impossible to sort on surname

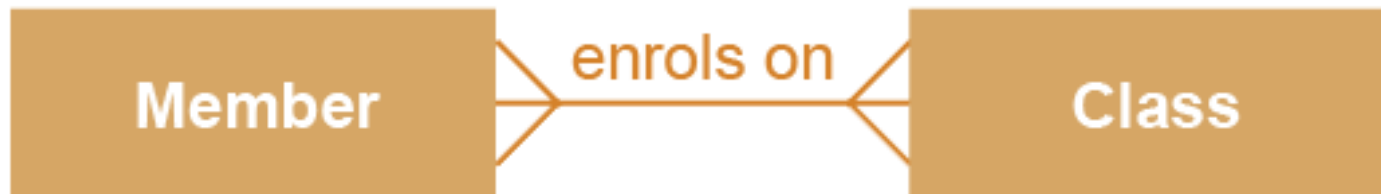
First normal form (1NF)

- Is the table below in first normal form?

MemberID	MSurname	MFirstName	ClassID	Cname	Day
M12	Sharpe	Norman	C100 C103 C114	Pilates Zumba Yoga	Monday Monday Wed
M13	Bunn	Martha	C100 C115	Pilates Cycling	Monday Wednesday
M16	Jones	Polly	C103	Zumba	Monday
M18	Jolly	Brian	C114	Yoga	Wed

First normal form (1NF)

- **No** – the table contains repeating groups of attributes
- the problem here is that the entities Member and Class are linked in a many-to-many relationship



Three tables in 1NF

- An extra table in the middle is needed



tblMemb

MemberID	MSurname	MFirstName
M12	Sharpe	Norman
M13	Bunn	Martha
M16	Jones	Polly
M18	Jolly	Brian

tblEnrolmen

MemberID	ClassID
M12	C100
M12	C103
M12	C114
M13	C100
M13	C115

tblClass

ClassID	CName	Day
C100	Pilates	Mon
C103	Zumba	Mon
C114	Yoga	Wed
C115	Cycling	Wed
C120	Fitness	Thur

Foreign keys

- By examining these three tables, can you identify the names of all members enrolled on the Monday Pilates class?
- What is the key of the Enrolment table?
- Are there any foreign keys in any of the tables?

tblMemb

MemberID	MSurname	MFirstName
M12	Sharpe	Norman
M13	Bunn	Martha
M16	Jones	Polly
M18	Jolly	Brian

tblEnrolmen

MemberID	ClassID
M12	C100
M12	C103
M12	C114
M13	C100
M13	C115

tblClass

ClassID	CName	Day
C100	Pilates	Mon
C103	Zumba	Mon
C114	Yoga	Wed
C115	Cycling	Wed
C120	Fitness	Thu



Composite primary key

- The key of the Enrolment table consists of the two fields MemberId and ClassID
 - It is a composite primary key
 - Both of these fields are also foreign keys
 - The tables are now in first normal form (1NF)

tblMemb

MemberID	MSurname	MFirstName
M12	Sharpe	Norman
M13	Bunn	Martha
M16	Jones	Polly
M18	Jolly	Brian

tblEnrolmen

MemberID	ClassID
M12	C100
M12	C103
M12	C114
M13	C100
M13	C115

tblClass

ClassID	CName	Day
C100	Pilates	Mon
C103	Zumba	Mon
C114	Yoga	Wed
C115	Cycling	Wed
C120	Fitness	Thu



Second normal form

- A table is in second normal form (2NF) if it is in first normal form and contains no partial dependencies
- This can only occur if the primary key is a **composite key**
 - **tblEnrolment** is the only table with a composite key and it has no fields which are dependent on either part of the key
 - The tables are therefore in second normal form

tblEnrolment

MemberID	ClassID
M12	C100
M12	C103
M12	C114
M13	C100
M13	C115

Third normal form (3NF)

- A table is in third normal form if it is in second normal form and contains no non-key dependencies
- This can be defined by saying:

*“All attributes are dependent on the key,
the whole key and nothing but the key”*

Third normal form

- To illustrate this, we will assume that the tblClass has extra fields specifying the ID and name of the instructor
 - In the revised table, which attribute do **InstFirstname** and **InstSurname** depend on?

tblClass

ClassID	CName	Day	InstID	InstFirstname	InstSurname
C100	Pilates	Mon	IN56	Abbie	Milton
C103	Zumba	Mon	IN67	Bob	Kelly
C114	Yoga	Wed	IN67	Bob	Kelly
C115	Cycling	Wed	IN56	Abbie	Milton
C120	Fitness	Thur	IN34	Jumal	Khan



Non-key dependency

- This is what is meant by a non-key dependency
 - There are four entities in this database, and each entity should have its own table



ClassID	CName	Day	InstID	InstFirstname	InstSurname
C100	Pilates	Mon	IN56	Abbie	Milton
C103	Zumba	Mon	IN67	Bob	Kelly
C114	Yoga	Wed	IN67	Bob	Kelly
C115	Cycling	Wed	IN56	Abbie	Milton
C120	Fitness	Thur	IN34	Jumal	Khan



Database tables in 3NF

tblMemb

MemberID	MSurname	MFirstNa me
M12	Sharpe	Norman
M13	Bunn	Martha
M16	Jones	Polly
M18	Jolly	Brian

tblInstructor

Instl D	Instfirstna me	instSurna me
IN34	Jumal	Khan
IN56	Abbie	Milton
IN67	Bob	Kelly

tblClass

Classl D	CName	Day	Instl D
C100	Pilates	Mon	IN56
C103	Zumba	Mon	IN67
C114	Yoga	Wed	IN67
C115	Cycling	Wed	IN56
C120	Fitness	Thur	IN34

tblEnrolmen

Memberl D	Classl D
M12	C100
M12	C103
M12	C114
M13	C100
M13	C115

Worksheet 3

Try the exercises on the worksheet



Importance of normalisation

- The advantages of normalisation are that:
 - It is easier to maintain and change a normalised database
 - There is no unnecessary duplication of data
 - Data integrity is maintained – if a person changes address, for example, the update needs to be made only once to a single table
 - Having smaller tables with fewer fields means faster searches and savings in storage

Importance of normalisation

- It is easier to maintain and change a normalised database

tblClass

ClassID	CName	Day	InstructorID
C100	Pilates	Mon	IN56
C103	Zumba	Mon	IN67
C114	Yoga	Wed	IN67
C115	Cycling	Wed	IN56
C120	Fitness	Thur	IN34

If we want to add an extra field to show the duration of each class, only one table needs to be changed

- Smaller tables with fewer fields means faster sorting and saves space

tblInstructor

InstructorID	Instfirstname	instSurname
IN34	Jumal	Khan
IN56	Abbie	Milton
IN67	Bob	Kelly



Importance of normalisation

- Data integrity is maintained

tblClass

ClassID	ClassName	Day	InstructorID
C100	Pilates	Mon	IN56
C103	Zumba	Mon	IN67
C114	Yoga	Wed	IN67
C115	Cycling	Wed	IN56
C120	Fitness	Thur	IN34

If Jumal Khan leaves, the database software will not allow the record to be deleted if the primary key is an attribute in a linked table. A new instructor first has to be allocated to class C120

Importance of normalisation

- Data integrity – if a person changes address, for example, the update needs to be made only once to a single table

tblInstructor

InstID	Instfirstname	instSurname
IN34	Jumal	Khan
IN56	Abbie	Milton
IN67	Bob	Kelly

If Abbie gets married and changes her surname to Fraser, the change is made in a single table

- “Data integrity” means that there is no possibility of having two different addresses (or any other attribute) for a person or item in the database

Plenary

- Do you know:
 - how to normalise relations to third normal form?
 - what properties are possessed by a relation in third normal form?
- Can you state:
 - Reasons why databases are normalised?
- Then you've cracked it!

Copyright

© 2016 PG Online Limited

The contents of this unit are protected by copyright.

This unit and all the worksheets, PowerPoint presentations, teaching guides and other associated files distributed with it are supplied to you by PG Online Limited under licence and may be used and copied by you only in accordance with the terms of the licence. Except as expressly permitted by the licence, no part of the materials distributed with this unit may be used, reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic or otherwise, without the prior written permission of PG Online Limited.

Licence agreement

This is a legal agreement between you, the end user, and PG Online Limited. This unit and all the worksheets, PowerPoint presentations, teaching guides and other associated files distributed with it is licensed, not sold, to you by PG Online Limited for use under the terms of the licence.

The materials distributed with this unit may be freely copied and used by members of a single institution on a single site only. You are not permitted to share in any way any of the materials or part of the materials with any third party, including users on another site or individuals who are members of a separate institution. You acknowledge that the materials must remain with you, the licencing institution, and no part of the materials may be transferred to another institution. You also agree not to procure, authorise, encourage, facilitate or enable any third party to reproduce these materials in whole or in part without the prior permission of PG Online Limited.